DOCUMENT RESUME

ED 393 261 EC 304 705

AUTHOR	Morison, Patricia, Ed.; And Others
TITLE	The Use of IQ Tests in Special Education Decision Making and Planning. Summary of Two Workshops. Board Bulletin.
INSTITUTION	National Academy of Sciences - National Research Council, Washington, DC. Commission on Behavioral and Social Sciences and Education.
SPONS AGENCY	Employment and Training Administration (DOL), Washington, D.C.
PUB DATE	96
NOTE	53p.; Summary report of two workshops of the Board on Testing and Assessment (LaJolla, CA, January 21, 1995 and Washington, DC, May 11, 1995).
AVAILABLE FROM	Board on Testing and Assessment, National Research Council, 2101 Constitution Ave., N.W., Washington, DC 20418.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Information Analyses (070) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Cognitive Measurement; Court Litigation; Curriculum Based Assessment; Decision Making; *Disabilities; *Disability Identification; Educational Diagnosis; Educational Planning; Educational Policy; Elementary Secondary Education; Eligibility; *Intelligence Quotient; *Intelligence Tests; Learning Disabilities; Mild Mental Retardation; Reading Difficulties; Referral; Special Education; *Student Evaluation; *Student Placement; Test Bias; Test Validity
IDENTIFIERS	Disproportionate Representation (Spec Educ)

ABSTRACT

This report synthesizes the key themes and ideas discussed at two 1-day workshops addressing issues in the use of IQ tests in special education evaluation and placement. The first section considers the background and legal context of the issue. The next section looks at IQ tests in school settings. Key ideas raised include the importance of never using a single test score to make decisions, inherent limitations of IQ scores, concerns about the validity of measuring complex mental abilities, and the relevance of the particular use to which the test is being put. A section on student placement considers the role of the IQ test at the stages of referral, evaluation, and placement. Alternative approaches to eligibility assessment, including curriculum-based measurement and assessment of cognitive correlates of reading failure, are reviewed. The need for further research in such areas as effective treatment methods, effects of various kinds of classroom environments on individual differences in learning, and the malleability of intelligence is stressed. Inset boxes provide summary information on such topics as the meaning and interpretation of IQ tests, IQ tests and the diagnosis of mild mental retardation, minority overrepresentation in special education, and use of IQ tests to diagnose learning disabilities. An appendix gives the workshop agendas and lists of participants. (Contains 68 references.) (DB)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BOARD BULLETIN

ED 393 261

The Use of IQ Tests in Special Education Decision Making and Planning

Summary of Two Workshops

EC304705

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
K. Barranco
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NATIONAL RESEARCH COUNCIL

BOARD BULLETIN

The Use of IQ Tests in Special Education Decision Making and Planning

Summary of Two Workshops

Patricia Morison, Sheldon H. White, and Michael J. Feuer, Editors

Board on Testing and Assessment
Commission on Behavioral and Social Sciences and Education
National Research Council

National Academy Press
Washington, D.C. 1996

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Harold Liebowitz is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Harold Liebowitz are chairman and vice chairman, respectively, of the National Research Council.

The work of the Board on Testing and Assessment is supported by the U.S. Departments of Defense, Education, and Labor, through a grant administered by the Employment and Training Administration of the U.S. Department of Labor.

Additional copies of this report are available from:
Board on Testing and Assessment
National Research Council
2101 Constitution Avenue N.W.
Washington, D.C. 20418

Copyright 1996 by the National Academy of Sciences. All rights reserved.
Printed in the United States of America

BOARD ON TESTING AND ASSESSMENT

Richard C. Atkinson (*Chair*), University of California, Oakland

Constance B. Newman (*Vice Chair*), Smithsonian Institution, Washington, D.C.

Richard J. Shavelson (*Vice Chair*), School of Education, Stanford University

Laurie J. Bassi, American Society for Training and Development, Alexandria, Va.

David C. Berliner, College of Education, Arizona State University, Tempe

Richard F. Elmore, Graduate School of Education, Harvard University

Patricia M. Flynn, Graduate School of Business, Bentley College

Edmund W. Gordon, Department of Psychology, City University of New York

Sylvia T. Johnson, School of Education, Howard University

Brigitte Jordan, Xerox Palo Alto Research Center and Institute for Research on Learning, Palo Alto, Calif.

Carl F. Kaestle, Department of Education, University of Chicago

Michael W. Kirst, School of Education, Stanford University

Luis M. Laosa, Educational Testing Service, Princeton, N.J.

Renée S. Lerche, Ford Motor Company, Dearborn, Mich.

Alan M. Lesgold, Learning Research and Development Center, University of Pittsburgh

Robert L. Linn, School of Education, University of Colorado, Boulder

Miles A. Myers, National Council of Teachers of English, Urbana, Ill.

James L. Outtz, Outtz and Associates, Washington, D.C.

Neal W. Schmitt, Department of Psychology, Michigan State University, East Lansing

Alan H. Schoenfeld, School of Education, University of California, Berkeley

William L. Taylor, Attorney at Law, Washington, D.C.

Ewart A.C. Thomas, Department of Psychology, Stanford University

Michael J. Feuer, *Director*

Patricia Morison, *Senior Research Associate*

Holly Wells, *Administrative Assistant*

Adrienne F. Carrington, *Administrative Assistant*

Membership as of the dates of the workshops summarized in this report. For current membership, see back cover.

PUBLICATIONS
Board on Testing and Assessment

Title I Testing and Assessment: Challenging Standards for Disadvantaged Children.
Summary of a Workshop

A Valedictory: Reflections on 60 Years in Educational Testing
By Lee J. Cronbach. Includes the report of the select committee on the California Learning Assessment System

Evaluation of the U.S. Employment Service Workplan for the GATB Improvement Project

Anticipating Goals 2000: Standards, Assessment, and Public Policy
Report of a Workshop

CONTENTS

Preface	vii
Background and Legal Context	1
IQ Tests in School Settings	5
Placing Children in Special Education	7
Alternative Approaches to Eligibility Assessment	19
Future Research	26
References	28
Appendix: Workshop Agendas and Participants	33

BOXES

1: The Meaning and Interpretation of IQ Tests	8
2: IQ Tests and the Diagnosis of Mild Mental Retardation	12
3: Minority Overrepresentation in Special Education	15
4: Using IQ Tests to Diagnose Learning Disabilities: Research and Policy Considerations	20
5: Curriculum-Based Measurement: An Alternative Approach	22
6: Cognitive Assessment of Reading Disability: An Alternative Approach	24

PREFACE

Intelligence testing, often heralded as one of modern psychology's most important contributions to human affairs, has been controversial since its invention at the turn of the century. Uncertainty and confusion over the many meanings of "intelligence," the validity of the intelligence test as a tool for estimating human cognitive ability, and the effects of using test scores as the basis for important decisions about people and organizations have given rise to periodic debates over the utility of testing in public policy. Although mental measurement can provide important insights to human performance, the validity of inferences about the origins, malleability, and implications of the measures is often challenged in public and legal forums.

One of the most controversial uses of intelligence (IQ) tests has been in decisions about the placement of children in special education programs. If, on the surface, the concept of identifying children with special learning needs and placing them in suitable programs is intuitively appealing, especially in a large democratic system committed to mass access to education, its implementation has always been fraught with conflict: the combination of evidence that disproportionate numbers of minority children were being classified for these special programs, with evidence that the programs were in many ways not providing adequate educational opportunities, has triggered complex and protracted legal actions. The landmark case, *Larry P. v. Riles*, resulted in the prohibition of the use of IQ tests with African American children in California; although this case is now more than 20 years old, the issues it raised are still current and at the center of another case now being tried in the California courts.

In the light of this renewed public interest in the uses of IQ tests in classifying children for special education, and particularly in the effects of test use on minority children, the U.S. Department of Education is attempting to forge a position that reflects its twin concerns—protection of civil rights and enhancement of educational opportunity for disabled students. To assist the department, the Board on Testing and Assessment convened two workshops to facilitate an examination and discussion of research evidence regarding the uses of IQ tests in special education placement decisions, with particular focus on mental retardation and learning disabilities. The workshops had the following objectives:

1. to provide an overview of legal, policy, and measurement issues regarding the use of IQ tests in special education;

2. to examine issues related to the validity and fairness of IQ testing for classification and placement of students in special programs, with emphasis on potential adverse effects on minority students; and

3. to explore some possible alternative assessment methods that could be used in combination with—or as substitutes for—traditional IQ tests.

This was clearly a big agenda for two 1-day workshops. Ten commissioned papers provided the basis for a lively discussion among a diverse group of scholars and policy makers.

This summary report is an attempt to synthesize the key themes and ideas discussed at the workshops; it is not a comprehensive, in-depth treatment of the research evidence on this complex topic. Many related issues, particularly on such a multifaceted topic, were barely touched on during the two days of discussion. However, the board is keenly interested in a set of issues related to the validity, fairness, and public understanding of testing that is broader than those that can be covered in this report; we hope to continue working on these issues in the future.

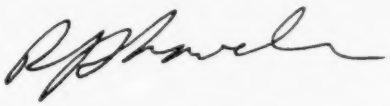
The workshop participants included board members and staff, officials of the U.S. Departments of Education and Labor, scholars from academic departments of law, psychology, special education, educational psychology, and public policy (for a list, see the Appendix to this report).

The work of the board is made possible through generous grants from the U.S. Departments of Defense, Education, and Labor. The continued interest and encouragement offered by Jane Arabian, Donna Dye, Emerson Elliott, Alan Ginsburg, Robert Litman, Gary Phillips, Valena Plisko, Elois Scott, Steve Sellman, Sharif Shakrani, and Raymon Uhalde are very much appreciated. These workshops would not have been possible without the assistance of several individuals at the Office for Civil Rights, the Office of Special Education Programs, and the Office of the Undersecretary, in the U.S. Department of Education: Sue Bowers, Art Coleman, Lou Danielson, and Jessica Levin provided thoughtful questions, comments, and very helpful support throughout the project.

Special thanks are due to Sheldon White, who so ably chaired both workshops, and to board members Robert Linn, Arthur Goldberger, and Luis Laosa for helpful comments on earlier drafts. We are also grateful to Milton Budoff, William Buss, Richard Durán, Stephen Elliott, James Greeno, Cynthia Hudley-Paul, Donald MacMillan, and Deborah Speece, as well as to board members David Berliner and William Taylor, for serving as discussants at the workshops. We are particularly grateful to senior research associate

Patricia Morison and board director Michael Feuer, who with Sheldon White took on the drafting of this summary.

Finally, thanks to Christine McShane, for her editorial skill and patience, and to Adrienne Carrington, who provided excellent administrative support, with able assistance from Holly Wells.



Richard J. Shavelson, Chair
Board on Testing and Assessment

The Use of IQ Tests in Special Education Decision Making and Planning

The Use of IQ Tests in Special Education Decision Making and Planning

BACKGROUND AND LEGAL CONTEXT

Few inventions of the social sciences have sparked as long or often bitter controversy as the "intelligence quotient." Although no longer calculated this way, IQ was originally derived as the ratio of an individual's mental age—as measured on one of a number of available standardized tests of intelligence—to his or her chronological age. The original intelligence test, developed by the French psychologist Alfred Binet at the turn of the century to help identify schoolchildren who were unlikely to benefit from regular instruction, has evolved into a sophisticated technology with individual diagnostic as well as large-scale assessment applications. Controversy over IQ, which has been chronicled by psychologists (Cronbach, 1975), historians (Fancher, 1985), educators (Chapman, 1988), media analysts (Snyderman and Rothman, 1988), and even natural scientists (Gould, 1981), can be attributed in large part to the difficulties in defining intelligence (Ceci, 1990; Feuer, 1994) and to widespread misconceptions over what intelligence tests really measure. As the National Research Council reported in 1982, the use of the word *intelligence* is "too easily misunderstood . . . [as] a unitary ability, fixed in amount, unchanged over time, and for which individuals can be ranked on a single scale" (Wigdor and Garner, 1982:28).

Although mental measurement has been heralded as one of psychology's most important contributions to human affairs (Cronbach, 1975; Herrnstein, 1971), controversy over the concept of intelligence, its origins, its distribution in the population, its malleability or mutability over time, and its role in determining social and economic opportunity has periodically exploded into heated public and political debate. The use of IQ tests to classify schoolchildren as learning disabled or educably mentally retarded—on the surface an application directly traceable to Binet's

original intent—has sparked an especially bitter and protracted battle involving the schools, parents, and, ultimately, the courts. Some legal battles, in some cases over 20 years old and involving people who have long since outgrown the age for which school placements would be relevant, are still simmering.

The dispute over special education classification originally arose from two fundamental issues: first, because students labeled as mentally retarded were often placed in separate special day classes for the educably mentally retarded (EMR) with a different, less academic curriculum that stressed social skills, leisure activities, and prevocational and vocational skills; and, second, because disproportionate numbers of the children classified as EMR were African American and Hispanic. IQ tests were implicated because they were used in the process of determining whether a child was to be classified as EMR.

In 1972, these issues came to a head in a now-famous lawsuit, *Larry P. v. Riles*, in which African American parents sought protection from the use of IQ tests to assess and place their children in California's EMR programs. Following an extensive trial in which the judge characterized the separate EMR classes as "dead end" and "inferior," the District Court in 1979 issued an injunction banning the use of IQ tests without prior court approval for the placement of African American students in EMR classes or their "substantial equivalent." In the years following, California abolished the separate category of EMR altogether. In 1986, the California Department of Education and the *Larry P.* plaintiffs agreed to a settlement that expanded the 1979 injunction to prohibit IQ testing for any African American student referred for special education eligibility evaluation.

The court in the *Larry P.* case did not focus heavily on examining the curriculum or the effectiveness of the special EMR classes. At that time, California had no evaluation data on the impact of its separate EMR classes and therefore no data by which to defend the placement of EMR children in special programs. Instead the court scrutinized the IQ tests themselves—whether they were valid for making decisions about EMR and whether they were racially biased. In its decision, the court ruled that IQ was "primary and determinative" and possibly the sole basis for labeling children as mentally retarded. The defense failed to convince the courts that other measures of adaptive behavior were influential in classification decisions. Furthermore, the court found that IQ tests were biased against minorities and thus resulted in disproportionate misclassification. There have been other significant court cases regarding IQ tests and

special education that have influenced educational testing policy, sometimes in contradictory ways; see Buss (1995) and Reschly (1995) for further discussion of legal issues.

However, the finding of bias in the use of IQ tests was fundamentally dependent on a hidden assumption of inefficacy of treatment. The judge decided that separate classes for EMR students were dead ends because students in them fell further and further behind and tended to remain in separate classes until the end of their school careers. *Had placement in these separate classes been perceived as beneficial, then overrepresentation of minority children might not have been viewed as a problem by parents, the public, or the courts.* "These negative views of the services offered in EMR classes are in marked contrast to the more positive perceptions of other programs designed to provide special services. For example, the significant overrepresentation of minorities in Title I programs has not been contested in major court cases, presumably because such children are *perceived* as obtaining effective remedial services designed to help them achieve the levels attained by their regular classroom peers" (Heller et al., 1982:5).

Despite the court-ordered ban in California, IQ tests are widely used to determine eligibility for special education. Many states and districts require the use of IQ tests because they are considered essential to the diagnosis of specific learning disabilities (SLD) as well as mental retardation. More recently in California, the problems of IQ testing of minority schoolchildren took a different turn when a group of parents of African American students having academic difficulties wanted their children to qualify for special education services as learning disabled. They filed suit, claiming that the ban on IQ testing violated their due process and equal access protections under the 1964 Civil Rights Act and the 1973 Individuals with Disabilities Education Act (IDEA). In this case, *Crawford v. Honig*, the plaintiffs alleged that they were being discriminated against on the basis of race and that the 1986 injunction violated the IDEA requirement that tests used for purposes of evaluation and placement of children with disabilities should be neither racially nor culturally discriminatory. They argued that African American students do not have access to tests that are available to other children, especially those used to qualify children as learning disabled.

The *Crawford* case was heard by Judge Peckham, the same judge as in the original *Larry P.* case, who vacated the 1986 injunction and ordered further proceedings. The Ninth Circuit Court of Appeals upheld Judge Peckham's ruling, remanding the case to the district court for proceedings (1) to determine "the contemporary meaning of the 'substantial equivalent' language" referring to any special education program(s) replacing the now defunct EMR classes and (2) to determine whether there are any effective uses of IQ

tests to make eligibility decisions about African American students for special education services.

Reflecting its twin concerns—protection of civil rights and enhancement of educational opportunity for students with disabilities—the U.S. Department of Education asked the National Research Council's Board on Testing and Assessment (BOTA) to facilitate an examination of the research evidence regarding the uses of IQ tests in special education decision making, with particular focus on mental retardation and learning disabilities.

Two workshops were convened with the following objectives: (1) to provide an overview of legal, policy, and measurement issues regarding the use of IQ tests in special education; (2) to examine issues related to the validity and fairness of IQ testing for classification and placement of students in special programs, with emphasis on potential adverse effects on minority students; and (3) to explore some possible alternative assessment methods that could be used in combination with—or as substitutes for—traditional IQ tests.

The first workshop, held in January 1995, focused on the current legal and policy context of IQ testing and special education, the validity of IQ as a measure of cognitive ability, and issues of diagnosis, classification, and placement in school systems. At the second workshop, held in May 1995, the focus was on alternative methods of assessment: what is known about alternative measurement technologies that can be used in conjunction with or as substitutes for IQ tests to classify children as learning disabled or mentally retarded. For each workshop, papers were commissioned from leading scholars with expertise in key topics (see the Appendix for the workshop agendas and participants).

This report summarizes the key themes and ideas presented in the papers and the discussion at the workshops, but with an important caveat: many of the topics discussed are extremely complex, and much of the related research evidence could not be covered in the allotted two days. Topics such as the relation between heredity and environment in the development of cognitive ability and its effects on achievement could not be covered. This report is therefore a synthesis of the workshops' key ideas and is not meant to be an exhaustive treatment of all the relevant topics. A companion volume, consisting of the papers presented at the workshops, will be published at a later date.

IQ TESTS IN SCHOOL SETTINGS

Although not nearly so widespread as was once the case, IQ tests are still used in schools, particularly tests that are individually administered by trained psychologists under prescribed, standardized conditions.[1] Most of these tests provide a range of subscores that represent relative strengths in different kinds of mental ability thought to make up general intelligence. A profile is obtained of a child's various abilities; such profiles can give the examiner information about patterns of abilities, strengths, and weaknesses.

IQ test score profiles are often used, in combination with a battery of other measures, to make decisions about individual children. Today their most widespread use in schools is to assist in determining eligibility for special education services through the Individuals with Disabilities Education Act, particularly for two of the most frequently occurring disabilities: mental retardation (MR) and specific learning disabilities (SLD). In most state education codes for the implementation of IDEA, the criteria used to define MR and SLD include IQ test scores. In 1992-93, 2.25 million SLD children and 468,000 MR children between the ages of 6 and 17 received such special education services; together these two groups account for 62 percent of all children receiving services (U.S. Department of Education, 1994).

An important maxim of appropriate test use is that no single test score should be used to make decisions about individuals. In special education, this idea is written into the IDEA legislation, which includes a number of provisions designed to protect students and ensure fair, nondiscriminatory use of tests. Among the provisions are that decisions about children must be based on more than a single test, that tests must be validated for the purpose for which they are used, that students must be assessed in all areas related to a suspected disability, and that evaluations should be made by a multidisciplinary team. Multiple sources of data should be used in decision making, including aptitude and achievement tests, teacher observations, parent interviews, health and medical information, and

[1]Discussion in the remainder of this report is limited to individually administered IQ tests. The dominant individual IQ test used with children is the Wechsler Intelligence Scales (WPPSI; WISC-III), although there are a number of others, such as the Stanford Binet Intelligence Scale and the Kaufman Assessment Battery for Children. Individual tests are personal; during a session "the tester gets to know the examinee, interacts so as to encourage a good performance, observes how the person approaches tasks and responds to frustration, and usually prepares a personalized interpretation" (Cronbach, 1990: 243). Group tests of intelligence, which allow many children to be tested at once, are rarely used to screen or refer children for special education and have not been at issue in any of the federal court cases dealing with disproportionate special education placement.

BEST COPY AVAILABLE

adaptive behavior. Nevertheless, some critics argue that due process protections requiring a battery of tests are not sufficient to prevent IQ tests from exerting undue influence in decision making and from "swamping" other evaluative information and test data.

What are some of the main limitations of individual IQ test results? First, test scores give only an estimate, at one point in time, of the level of an individual's performance; test results tell little about the processes by which the child reached that level. Second, tests of ability assume maximum performance—that the student being tested is trying to do his or her best, that each understands what is expected and that each is willing to comply with the examiner and the testing directions (Snow, 1995). Yet these important caveats about the meaning and interpretation of test scores are not well understood by many educators, parents, students, or the general public. Nor have the assumptions about maximum performance been empirically evaluated, although considerable research attention has been paid to the importance of considering the testing process and the effects of interactions between examiner and child on test performance (see, for example, Sattler, 1982; Salvia and Ysseldyke, 1985).

The validity question for tests that purport to measure complex mental abilities, like intelligence, is a multifaceted one. Multiple sources of evidence must be considered (see Box 1). Different sources of evidence must be woven together to support a judgment about the meaning and worth of a particular test application. A presentation at the workshop proposed that validation also must actively consider counterarguments and alternative hypotheses—which are often not a primary concern in test development. One such counterargument, long raised as a challenge to the validity of IQ tests, is that these tests are racially or culturally biased. An extensive literature exists regarding the several meanings of test bias; this literature was not summarized at the workshops; see Wigdor and Garner (1982), Cole and Moss (1989), and Holland and Wainer (1993), for further discussion. Similarly, evidence regarding the validity of tests for people with various disabilities was not reviewed; see, for example, Sherman and Robinson (1982), and Willingham et al. (1988). Some argue that the value implications and social consequences of test interpretation and use also need to be considered as part of the validity argument—a relatively new psychometric idea (Snow, 1995).

Participants at the workshop suggested that the courts have ultimately tended to think about validity for IQ tests with great emphasis on the consequences of the particular use to which the test is being put. In K-12 educational settings, the purpose of a test, as well as the criteria against which it will be validated, is often much less clear than, for example, criteria for validating employment or college admissions tests. For IQ tests

(or any test that purports to measure a general ability or G), some researchers emphasize that it is important to look at how well G predicts performance in different educational settings; they argue that cognitive ability (or intelligence) is a function of the situation as well as the person. This view is distinctly different from the predominantly held notion of IQ (or intellectual ability) as a fixed, predetermined amount that somehow limits one's options and predicts future success (Snow, 1995).

There is some evidence that IQ measures can predict how children respond to different kinds of instruction. Evidence presented at the workshop indicates that, across the whole range of children in schools, "G measures relate more strongly to learning outcome when instruction can be characterized as relatively incomplete or unstructured, involving discovery or inductive teaching methods, and placing heavy demands on learner self-regulation, information processing and inferential reasoning" (Snow, 1995:16).[2] Further research of this type, relating assessments to response to instruction, could lead to better design of instruction that fits student cognitive strengths and weaknesses. Currently, however, we do not have validity evidence regarding the capacity of IQ tests to predict how children with specific IQ test score profiles will respond to different kinds of special education treatments.

PLACING CHILDREN IN SPECIAL EDUCATION

As IDEA is implemented in schools today, there are generally three stages that children pass through in the process of receiving a disability diagnosis and special education services.

Stage One: Referral

Often overlooked in discussions of which children get labeled and placed in special education is the referral process. Most children are never screened or tested to discern whether or not they should be considered for special education. Instead, referral (usually by teachers) is the primary method through which children begin to be considered for the disabil-

[2] As this finding illustrates, research with IQ measures has contributed greatly to our understanding of the complex development of intellectual skills across the whole population. However, the usefulness of a particular test as a basis for planning or predicting how an individual will do is much more limited.

BOX 1
The Meaning and Interpretation of IQ Tests

Defining intelligence is fraught with conceptual and ideological complexity. Tests that are intended to estimate the construct of intelligence, general ability, or G are beset with additional technical and conceptual difficulties. As Snow remarked (1995:7):

> Unfortunately . . . simplistic "trait" interpretations of test scores and their intercorrelations have misled some scientists, and much of the public, into thinking of fundamental, fixed entities residing in the head and measured as physical amounts; terms such as "potential" and "capacity" are also often found in such interpretations, as if the score could foretell future limits on development. Again, it would be a major advance for both scientific and public debates on mental testing if the newer, complex process interpretation of ability could be made to replace the old, fixed entity interpretation in common parlance.

Essential to the usefulness of any test of intelligence is validity evidence of many kinds. In his paper, Snow described 10 categories of validity evidence that are needed for cognitive ability tests:

1. Convergent and discriminant correlations: evidence that different tests measuring G correlate highly with one another and less strongly with tests of other, unrelated constructs.

2. Predictions of performance criteria: evidence that G measures predict later performance on related criteria, such as educational attainment and academic grades.

3. Prediction generalization and differentiation: evidence regarding prediction as a function of different treatment conditions, such as different kinds of educational programs.

4. Content and process analyses: evidence regarding how and why the content and process of performance on tests of G and on learning tasks are highly related in some situations but not others.

5. Boundaries on interpretation and use: identification of the boundaries within which specified interpretations and uses are justified and outside which they are not.

6. Short-term malleability: demonstrations of the degree to which scores are subject to short-term changes and by what means.

7. Long-term development: examination of long-term change, both as a result of intervention and of normal development.

8. The adequacy of measurement models and methodology: evidence regarding the adequacy of existing measurement models especially regarding technologies for diagnostic measurement and for detecting bias.

9. The adequacy of selection, classification, and education decision rules: evidence regarding the decision rules and cut-off points for classifying students into different instructional treatments, such as special education, pull-out remedial instruction, and alternative special classes.

10. The consequences of the proposed interpretation and use, and its social value implications: full consideration of the consequences of test use will require a comprehensive evaluation of tests and special education treatments as a system.

In concluding his review, Snow remarked (p.34):

> Validation is a continuing process. The validity question for G tests may never be finally and fully decided, given likely perpetual changes in population and conditions of use. But the existence of G as a psychological phenomenon and the standards and measurement methods used for its assessment are quite well substantiated. To the extent that a given "IQ" test conforms to these standards and methods, including those covering evaluation of the particular proposed use, then one can claim provisional validity for this test interpretation and use as a measure of G.
>
> Validation is also a multifaceted process, however, and claims of valid interpretation and use in particular instances must marshal evidence concerning that instance from many relevant sources. Given both the continuous and the multifaceted nature of the problem, the best practice in any given instance would be to choose the best test for the purpose, use it strictly in accordance with professional standards, and evaluate its use in place with periodic, longitudinal, empirical study.

Based on a paper presented at the first workshop by Richard E. Snow (1995). For a more complete discussion of historical, legal, conceptual, and technical issues in ability testing see, for example, Wigdor and Garner (1982) and Cronbach (1990).

ity diagnosis and delivery of special education services. Medical disabilities (e.g., sensory or orthopedic impairments) are usually diagnosed long before a child reaches school age, do not rely on psychological assessment, and show similar rates of prevalence from place to place (e.g., from state to state). In contrast, children with other kinds of disabilities—those with no medical or physical anomalies—make up the majority of disabled children and typically have difficulties that are largely exhibited *within* the school system. Usually such students enter school unidentified and are served in the regular classroom until they begin to exhibit problems with achievement or, in rarer instances, behavior. "Teacher referral typically is the first and most important step in the process whereby students are diagnosed as mildly disabled, and the vast majority of students with mild disabilities will not be designated officially as disabled during their adult years" (Reschly, 1995:8). Although children can be referred by parents, doctors, social workers, or others, most are referred by their teachers because of "repeatedly poor academic performance or poor social adjustment" (Heller et al., 1982:38). Often those referring a child must meet with a study team to present evidence that the general classroom is not serving the child's learning needs. This team often requires evidence that various adaptations have been tried in the general classroom before a child can continue to the next stage of formal evaluation.

Stage Two: Evaluation

IDEA is explicit and detailed in its prescriptions regarding evaluation procedures. Children are generally tested and observed by several different professionals, such as the school psychologist, a special education resource teacher, a speech and language therapist, an occupational therapist, and a reading specialist. Administration of an IQ test is almost always required for any child being considered for either the MR or SLD diagnosis. School psychologists, who have special training in test administration and interpretation, are responsible for IQ testing in most states. This evaluation stage often takes at least several months and can be quite expensive. One estimate suggests that the cost of the typical eligibility assessment process is often higher than the reimbursement for educating a child with a mild disability for an entire year (Fuchs, 1995).

At the end of the evaluation, an eligibility meeting is held at which the evidence is shared and discussed among professionals and reviewed with parents. A decision is made about whether or not the child meets the requirements for any of the disability classifications.

The current classification system used in special education provided under IDEA is driven by the notion of a disability. Although there is no universally recognized official

classification system for children in educational settings, the regulations for IDEA offer brief definitions of 13 disabilities: autism, deaf-blindness, deafness, hearing impairment, mental retardation, multiple disabilities, orthopedic impairment, other health impairment, serious emotional disturbance, specific learning disability, speech or language impairment, traumatic brain injury, and visual impairment (Individuals with Disabilities Education Act, 1991; see 34 CR 300.7). However, to qualify for special education services, not only must the child be eligible for a disability designation, but he or she must also demonstrate a need for special education or related services. This two-pronged eligibility definition means that not all disabled children necessarily receive special education services. Thus, for example, children with medical or physical disabilities do not qualify to receive special education services unless they also demonstrate educational need.

States vary in the extent to which they adopt the IDEA categories and definitions; as a result, "it is entirely possible for students with identical characteristics to be diagnosed as disabled in one state, but not in another, or to have the categorical designation change with a move across state lines" (Reschly, 1995:5). Administrative policies can exert substantial influence on how children get classified in various categories, particularly the mild disabilities. For example, the percentage of children classified as SLD varies from 2.73 percent of the school-age population in Georgia to 9.43 percent in Massachusetts; MR varies from 0.33 percent in New Jersey to 3.06 percent in Alabama; speech or language impairment (SP/L) varies from 1.16 percent in Hawaii to 4.23 percent in New Jersey; and serious emotional disturbance (SED) varies from 0.04 percent in Mississippi to 2.08 percent in Connecticut (U.S. Department of Education, 1994). "It is highly unlikely that there are over nine times as many students with mental retardation in Alabama than New Jersey; that there are over three times as many students with SED in Connecticut as in Mississippi"(Reschly, 1995:10). Nationwide, 10.24 percent of children ages 6-17 receive a disability diagnosis; the states, however, vary from a low of 7.09 percent in Hawaii to 15.21 percent in Massachusetts, suggesting that state-by-state practices exert very real influences on prevalence rates for disability diagnoses.

When making nonmedical disability diagnoses, IQ tests are central to distinguishing between SLD and mild mental retardation (MMR). Although specific definitions and criteria vary, MMR is generally viewed as low achievement combined with lowered general intellectual ability (measured by IQ tests, see Box 2), whereas SLD is commonly viewed as "unexpected poor performance" or low achievement that is significantly lower than would be predicted for a child with normal or above-average general intellectual ability. The diagnosis of SLD is often sought by parents whose children are encountering aca-

BOX 2
IQ Tests and the Diagnosis of Mild Mental Retardation

IQ tests are particularly salient when trying to distinguish among three groups of children who exhibit "severe and persistent academic underachievement": mildly mentally retarded (MMR), specific learning disabled (SLD), and low-achieving (LA) students who fail to qualify as either MMR or SLD. The diagnosis for more severe forms of retardation—those caused by major chromosomal anomalies (e.g., Down syndrome), metabolic disorders (e.g., phenylketonuria), and other origins that are biological in nature—does not rely on IQ; "these children typically exhibit behaviors that are judged maladaptive by any standard and in any setting (e.g., school, home, community)" (MacMillan, 1995:2).

The controversy surrounding the use of IQ tests and the legal challenges to their use with minority children has focused on those cases of MR for which there is *no* biological etiology that can be determined, no physical stigmata evident, and obtained IQ scores that range from 60 to 90. To date the meaning of an IQ of 45 has not been challenged, regardless of the ethnicity or race of the child, when accompanied by a "general dysfunction" in the areas of self-help skills, independent functioning, and maladaptive behaviors.

The definition of mild mental retardation is not clear cut, but most professionals rely on the definition provided by the American Association of Mental Retardation (AAMR) (Luckasson, 1992). In defining MMR for policy purposes, states differ in their specific requirements but most include IQ; there is usually an upper IQ level specified (usually between 70 and 75, although some states such as Iowa go as high as 85). This differs considerably from conditions at the time of the original *Larry P.* hearing, when an IQ of 85 was the common upper limit. Far fewer children are identified today as MMR, and those who are exhibit more generalized dysfunction than was the case two decades ago.

In addition, however, since the early 1960s the various versions of the AAMR definition of mental retardation require more than just low cognitive ability—

all require a child to exhibit impairments in adaptive behavior (Luckasson, 1992). Although the construct of adaptive behavior has been difficult to measure, its use has served to prevent the vast majority of individuals with IQ scores below cutoffs from being labeled mentally retarded. Referral by teachers serves as the first and most comprehensive screen for such adaptive behavior. According to MacMillan, many children who, if they were to be tested, would score below an IQ cutoff for MMR are maintained in regular grades and complete their public schooling without ever taking a test of intelligence. In addition, it is the child's academic performance—reading, computing, spelling, and so on—that puts him or her at risk for referral—not IQ.

How influential are IQ test scores in making the diagnosis of MMR? Over the years the courts have reached quite different conclusions regarding this question. In the *Larry P.* case, Judge Peckham concluded that IQ was "primary and determinative," whereas in a similar case (*PASE v. Hannon*, 1980) the judge concluded that IQ was only "part of the process" through which children were classified as MMR (Reschly et al., 1988).

Data from studies of how referred children actually get placed (or not) in special education suggest that, in practice, IQ is not the "sole or primary basis of placement" (MacMillan et al., 1995). In addition, data suggest that IQ can serve as a two-edged sword in classifying children as MMR: on one hand, it can result in finding the child eligible when the IQ is *below* the cutoff score; on the other hand, when an IQ score *exceeds* the cutoff, it can prevent a child from being labeled and placed.

Additional data presented by MacMillan suggest that schools, at least in California, are extremely reluctant to assign the MMR label to students, even when those students meet the objective criteria of the AAMR-92 definition. These data suggest that the more acceptable label of SLD is often assigned to these children and may account (at least in part) for the increased prevalence of this diagnostic category and the decreased prevalence of MMR.

Based on a paper presented at the first workshop by Donald MacMillan (1995).

demic difficulties in the classroom and who hope to receive extra services for their children. Each of these is distinguished from simple "low achievement"—severe and persistent underachievement in children whose intellectual ability is neither high nor low enough to enable them to qualify for either the SLD or MMR diagnosis (MacMillan, 1995).

Since the *Larry P.* case (and perhaps because of it), there has been a large decline in the number of children who receive the MR label, accompanied by a large rise in the number of children being called SLD. Since 1976, when data were first collected on the prevalence of various disabilities, the prevalence of SLD has more than doubled from 23.8 percent of all students classified as disabled to 52.4 percent in 1992-93 school year. The prevalence of the MR diagnosis has declined since 1976 from 24.9 percent to 10.9 percent in 1992-93 (U.S. Department of Education, 1994).

Overrepresentation of African American children still occurs in certain categories of disability, although it has decreased in the past 20 years (see Box 3). Hispanic children are somewhat underrepresented in special education. Data from the 1990 Office for Civil Rights (OCR) survey suggest that, over the four categories of MMR, SLD, SP/L, and SED combined, about 11 percent of African American and Native American children receive special education services, compared with 9.5 percent of whites, 8 percent of Hispanics, and 4 percent of Asian/Pacific Islanders (U.S. Department of Education, 1994).

IQ tests are used in special education primarily as a classification tool to designate a child as belonging to a particular category of disability; policy guidelines often require their use. In the implementation of PL 94-142 at the state and local levels, IQ tests are often specified as one of the tools needed to diagnose SLD and MMR. In the case of MMR, there is often an upper IQ test score specified. In defining SLD, most policy guidelines include the presence of a significant discrepancy between a child's IQ test score and achievement scores—that is, a child is often considered SLD if achievement levels are significantly lower than his or her IQ test scores. Policy guidelines and discrepancy criteria that rely on standardized IQ tests represent an attempt to apply some kind of common metric across districts and school buildings to achieve reliable and consistent SLD and MMR diagnoses. Indeed, according to participants at the workshop, IQ testing remains largely a classification tool with limited application to the instructional planning process (stage 3). Because most IQ tests with good technical characteristics produce normative information about a child's relative standing in a group, they are useful for classification but not very useful for planning instruction. The Buros Mental Measurements Yearbook notes of the most widely used IQ test for

BEST COPY AVAILABLE

BOX 3
Minority Overrepresentation in Special Education

Minority overrepresentation in special education, especially for the nonmedical disability categories, has been a concern for several decades. In the years since 1972, when the *Larry P.* case was filed, there have been dramatic changes in the use of some of these education-based categories of disability. For example, since 1976, the number of children receiving the MR designation has dropped by more than half (from 24.9 percent of the total number of children classified as disabled to 10.9 percent in 1992-93), whereas the number of children classified as having a specific learning disability (SLD) more than doubled over that same time period, from 23.8 to 52.4 percent (U.S. Department of Education, 1994). (During this same time period the total number of children classified as disabled rose from 3.7 to 5.2 million.)

Patterns of minority representation in different categories have changed as well. Table 1 presents nationally representative data regarding the classification rates for African American, Hispanic, and white students in nonmedical (or "mild") disability categories. The table presents the only national data regarding overrepresentation, which are collected by the Office for Civil Rights (OCR) in the U.S. Department of Education. Several cautions must be made regarding these data, however. The OCR surveys are not a representative sample of school districts in the United States but rather include all of the "large" districts and a sample of "smaller" districts. In addition, the OCR sampling strategy has changed for the surveys reported in the table; it is not possible to determine the effects of these varied sampling strategies on the results (Reschly, 1995).

Several trends regarding minority participation in special education are apparent from these data. First, the base rates of identification as mildly disabled have increased for all three racial ethnic groups since 1978, the biggest proportional increases occurring for white children (from under 4 percent identified in 1978 to over 6 percent in 1990). In 1990, African American students still have the highest rate of representation (about 8 percent of African American children received one of the three mild disability diagnoses) compared with about 6.5 percent of white and 5.7 percent of Hispanic children. However, the overrepresentation difference between the proportion of African American children and white children identified has narrowed over this same time period, from about 2.5 percent in 1978 to 1.5 percent in 1990.

Patterns of disproportion vary by diagnosis. The majority of the identified children in all three racial ethnic groups receive the SLD diagnosis, and at approximately similar rates (just under 5 percent of all three groups are identified as SLD). Across all three categories of mild disability, Hispanic children show the lowest rates of identification. Most of the difference among the three racial ethnic groups is due to the larger percentage of African American children labeled MMR (2.1 percent compared with 0.7 percent of Hispanic and 0.8 percent of white children). African American children also show slightly higher rates of identification as SED.

A number of factors have been discussed as potential causes of overrepresentation, including legal and administrative requirements, characteristics of students, quality of the instruction received, possible biases in the assessment process, and characteristics of the home and family environment, as well as broader historical and cultural contexts (Heller et al., 1982). Data are needed, however, that would allow analysis of the relative effects of these factors on placement rates for various special education diagnostic categories.

In his paper, Reschly presented data from four school districts that allow the preliminary examination of the effect of poverty and minority group status on special education placement rates. Across these four districts, far more African American children (62 percent) qualified for subsidized school lunches (a proxy for poverty) than did white children (15 percent). (The overall school population was 31 percent African American, 65 percent white, and 4 percent other.) Those children, both white and African American, who qualified for subsidized school lunches had much higher rates of placement in special education (the median placement rate was 19 percent for both groups) than did their more advantaged peers (the range across districts was from 7.3 to 9.2 percent for African Americans and 5.3 to 7.3 percent for whites). Furthermore, in three of the four districts, African American and white children in poverty had equal rates of classification. Reschly concluded that "poverty circumstances are a plausible explanation for the special education overrepresentation, although additional studies are needed of the nature and magnitude of the relationship of poverty to disability diagnosis and special education placement" (Reschly, 1995:19).

Based on a paper presented at the first workshop by Daniel J. Reschly (1995). Although there are 13 categories of disability, the vast majority of classified children receive one of the nonmedical or "mild" disability diagnoses.

TABLE 1 Percentages of African American, Hispanic, and White Students Classified as Mildly Disabled

	Disability Category			
Year	MMR	SLD	SED	Total
1978:				
African American	3.46	2.23	0.50	6.19
Hispanic	0.98	2.58	0.29	3.85
White	1.07	2.32	0.29	3.68
1986:				
African American	2.30	4.43	1.04	7.77
Hispanic	0.56	4.31	0.46	5.33
White	0.87	4.29	0.57	5.73
1990:				
African American	2.10	4.95	0.89	7.94
Hispanic	0.65	4.68	0.33	5.66
White	0.81	4.97	0.69	6.47

Note: Data in the table represent the percentage of the total number of students in each ethnic minority group who are classified in the particular category given. MMR = mild mental retardation; SLD = specific learning disability; and SED = seriously emotionally disturbed.

Sources: 1978 data based on Finn (1982:324-330). 1986 data based on analyses by Reschly and Wilson (1990), using 1986 OCR survey data compiled by the National Council of Advocates for Students. 1990 data based on U.S. Department of Education (1994:198, 201, 202).

children ". . . the WISC-R lacks treatment validity in that its use does not enhance remedial interventions for children who show specific academic skill deficiencies" (Witt and Gresham, 1985:1717).

Yet even the usefulness of IQ tests as a classification tool has been challenged. Papers presented at the workshop suggest that the criteria for making a diagnosis of MMR has changed and that IQ tests play a lesser role than they once did (see Box 2 and MacMillan, 1995; Reschly, 1995). There is also growing evidence that diagnosing SLD on the basis of a discrepancy between IQ and achievement test scores—largely a policy-based definition—has not held up under empirical scrutiny (see Box 4 and Fletcher et al., 1995).

Some of these findings call into question the diagnostic system as it is currently being implemented under IDEA. The words of an earlier National Research Council report, issued over a decade ago, remain remarkably relevant to this point (Heller et al., 1982:21):

> A significant question also arises as to the mechanism by which special instruction may best be provided. In particular, to what extent must children be classified and labeled according to a generic class of deficiencies in order to receive special education services? Diagnostic categories such as EMR may be more an administrative convenience than an educational necessity, allowing schools to count the number of children in this and other special programs in accord with federal requirements. If categorical labels are required for administrative purposes, they could be chosen to reflect the educational services provided, thereby emphasizing the responsibilities of school systems rather than the failings of the child.

Stage 3: Placement

Once a child is found eligible and receives a diagnosis, the professionals responsible for delivering the services write an individualized education program (IEP). The IEP specifies a plan for the kinds of instructional treatments the child should receive, including whether the child will be placed in regular or special classes, full or part time. Evidence suggests that the IEP has evolved largely into a procedural document; as an instructional planning device, it has fallen short of the original intent. It is often vague on educational goals, treatments, and procedures for evaluating those goals (Smith, 1990). In addition, evidence suggests that the placement recommended, as well as the IEP content, are often determined by the resources, classes, and teachers available, rather than the needs of the child (Heller et al., 1982). Legal, fiscal, and organizational constraints influence how children get referred, labeled, and placed. At the most basic level, the amount of money a school or district has available for special education, as well as the overall needs of the school's or district's population, exerts a pronounced

effect on who gets into the special education system and what kind of instructional services they receive (Mehan, 1995).

ALTERNATIVE APPROACHES TO ELIGIBILITY ASSESSMENT

Misclassification can result in children's receiving the wrong educational treatments. Indeed, the *Larry P.* case became a symbol of generalized public anxiety over testing and test use because of the underlying fear that children labeled mentally retarded were being placed in poor educational environments. Although tests may serve many functions, in education their primary diagnostic functions should be, in the view of the workshop participants, twofold: identifying students who can profit from special forms of instruction and specifying what kind of intervention or instruction will be effective. *Tests used during the special education eligibility process are not closely linked to decisions about instructional treatments*. Instead, assessment is primarily used as a classification tool. This is not a new finding, but rather a problem that has vexed policy makers for many years. In their 1982 report, for example, the National Research Council's Panel on Selection and Placement of Students in Programs for the Mentally Retarded drew attention to the role of testing as a tool to facilitate effective intervention (Heller et al., 1982:46):

> Our decision to focus on the educational utility of various assessment devices and procedures, rather than on their role in classification and misclassification, is based primarily on the fact that we are analyzing assessment in an educational context, in which it is a means to the end of improving instruction.

In recent years, diverse bodies of research have converged toward a deeper understanding of the role of context in learning, the development of cognition and individual differences, entry into school, and risk factors associated with later school failure. In addition, evidence has accrued regarding the effectiveness of some instructional methods with specific populations of disabled students (see, e.g., Keogh and MacMillan, 1995; Fuchs and Fuchs, 1995). These research efforts offer important data that can help shape new assessment strategies. In addition, alternative assessment tools, such as performance assessment for use in classrooms (see Elliott, 1995), have received considerable research and policy attention. Assessment aligned with instruction—an important vision in general education—also deserves a place in special education decision making (see, e.g., Fuchs, 1994). *Research focused on validating assessments hand in hand with effective treatments offers considerable promise for special education decision making.*

BOX 4
Using IQ Tests to Diagnose Learning Disabilities: Research and Policy Considerations

The definition of specific learning disabilities (SLD) embodied in PL 94-142 and the Individuals with Disabilities Education Act is based on obtaining IQ test scores that are significantly higher than achievement test scores. This definition of SLD has its origin in early medically based notions of brain dysfunction in children and its relation to learning problems. Early research on children with unexpected reading failure suggested that there were two distinct groups represented by a "hump" at the lower tail of the distribution of reading skills: one with a specific reading disability and others who were generally "slow learners" (Rutter and Yule, 1975). The children identified as having a specific reading disability were those who exhibited a discrepancy between actual reading performance and predictions of reading performance based on IQ scores—and thus were thought not really to "belong" at the lower tail of the distribution. In the United States, this latter group are generally considered eligible for special education services as SLD; the "slow learners" are not.

The discrepancy notion was a hypothesis when it was included in the *Federal Register* some 20 years ago. But, according to the presentation by Fletcher, recent research challenges its empirical validity. Data have been gathered to examine the validity of this "two-group" hypothesis regarding learning disabilities: that is, that a group of children identified on the basis of a signifi-

Two examples of research that links assessment and instruction—one focused on identifying cognitive correlates of a particular type of learning difficulty, the other on an assessment method sensitive to changes in a child's rate of learning or growth in different learning environments—were presented at the second workshop. These lines of research suggest new paradigms for the role of assessment in special education as well as methods for identifying and implementing effective instructional treatments. Researchers are exploring other assessment and instructional methods as well. For example, "dynamic assessment" and other methodologies that draw on the classic work of Vygotsky and other psychologists in the Soviet Union focus on how well the child learns in the testing situation. Instead of trying to elicit the "maximum performance" that a child can produce unaided, these methods provide coaching and help during the testing situation so that a child can exhibit his or her full ability to learn (see Lidz, 1987). An-

cant discrepancy between ability (assessed by an IQ test) and achievement is distinctly different (i.e., is learning disabled) from a group of children whose IQ and achievement scores are both low. Several studies of the validity of the two-group classification system yield similar findings: they provide insufficient evidence to support the hypothesis that children with an IQ-achievement discrepancy can be differentiated from children whose achievement levels are not discrepant with IQ (Fletcher et al., 1994; Foorman et al., 1995; Francis et al. 1995; Stanovich and Siegel, 1994). Furthermore, data suggest that early reading patterns and cognitive correlates of both groups are the same. Although the evidence suggests that the children who show a discrepancy do have a valid learning disability, it does not support the hypothesis that they have some special characteristics that separate them from children with academic deficits who fail to show a discrepancy. According to Fletcher, the data suggest that reading disability occurs on a continuum of normal reading capabilities.

A further serious limitation of using the discrepancy model is that it often postpones the identification of many children as SLD until the third or fourth grade (in part because it takes that long for their achievement scores to move sufficiently far from the floor of the test to obtain a discrepancy). This is particularly problematic because research evidence suggests that reading disabilities identified after age 8 are more resistant to treatment than those identified earlier (Strag, 1972; Muehl and Forell, 1973).

Based on a paper presented at the second workshop by Jack Fletcher et al. (1995).

other kind of assessment focuses on specifying a child's error patterns and misconceptions in a subject area (e.g., mathematics) and can lead directly to instruction focused on remediation of conceptual misunderstandings (Brown and Campione, 1986). These research programs focus on identifying effective instructional methods as well as specific assessment tools.

Curriculum-Based Measurement

Curriculum-based measurement (CBM) uses weekly and systematic but brief assessments of a child's progress toward expected learning outcomes in the regular curriculum (Fuchs, 1995; see also Box 5). By assessing children weekly using methods embedded in the instructional process, teachers can identify children who fail to make expected progress in learning. Various educational interventions can then be tried to see whether more or

BOX 5
Curriculum-Based Measurement:
An Alternative Approach

Curriculum based measurement (CBM) is a set of methods for indexing academic competence and progress. Based on a program of research conducted at multiple sites over the past 15 years (see, for example, Deno, 1985; Shinn, 1989), CBM specifies methods for sampling test items from local curricula (in spelling, math, and reading), for administering and scoring those assessments, and for summarizing and interpreting the resulting data. For example, two measures are used for reading: (1) reading aloud from passages (scored as number of words read correctly per minute) and (2) restoring blanks to passages (scored as number of correct restorations in 2.5 minutes). Teachers typically assess children weekly in a process that takes about 5 minutes per child (and can be done using computers). Each child then receives a graph that shows how much of the curriculum has been mastered, how much progress he or she is making over time, and how that compares with the progress of peers. Because the assessment is linked closely to the local curriculum, teachers can use the assessment information to modify instruction. Thus, according to Fuchs, CBM bridges traditional psychometric approaches with classroom-based observational assessment methods. Such bridging provides information about individual academic standing as well as growth, about global competence as well as specific skill mastery, and information about how to strengthen academic programs.

CBM was originally developed for two purposes: to monitor students' academic progress and to link instructional planning with assessment in order to enhance student outcomes. Although CBM was not designed as a tool for identifying students for special education, subsequent studies have suggested that CBM can differentiate school-identified students with specific learning disabilities (SLD) from non-SLD peers (e.g., Shinn et al., 1987). Because the comparison group, as well as the assessment tasks, are locally determined, CBM used in eligibility decisions results in a situational notion of disability. Problems are defined locally with respect to both the curriculum demands and peer group, offering the logistical advantage of allowing schools to serve those students with the greatest immediate academic needs. In the past 15 years, CBM has been used in a number of schools to formulate eligibility decisions in a 2-stage process consisting of problem identification and problem certification (e.g., Marston and Magnusson, 1988; Marston et al., 1984).

Based on a paper presented at the second workshop by Lynn Fuchs (1995).

better growth takes place. The CBM approach embodies an attempt to identify and address problems of learning while keeping the child in the regular classroom environment.

Although still in the pilot stages, a three-stage treatment validity approach was described for using CBM to make decisions about eligibility for special education. First, CBM would be used to determine whether an individual student's level of performance and rate of growth are comparable to that of other students in the same class. Has the child's rate of learning in reading, for example, kept pace with that of others in the class? If the growth rate and performance levels lag behind peers, then a second stage involves modifying the general education learning environment to see whether the student's learning rate can be increased. If not, the third stage would ask whether special education can effect better growth. For some period of time, the child would receive special education instruction to see whether better growth can be achieved. If so, then a positive eligibility decision would be made. This three-phase process has two key purposes: (1) to identify and correct problems wherever possible and (2) to make sure children can actually benefit from the services provided.

Although the effects of this approach to eligibility on the demographics of the identified population of children are not yet known, this treatment-oriented approach would be aimed at reducing inequities by: (1) maximizing the potential of regular education, (2) reserving judgment about the need for special education, and (3) verifying that special education is valuable and desirable for the child. The question of what to do with children who don't respond to special education treatments remains a difficult problem.

Assessment of Cognitive Correlates of Reading Failure

The second line of research presented at the workshops provides an example of a diagnostic role for assessment in special education—one that can help select the appropriate treatment based on characteristics of the child. In this method, a battery of tests are designed to assess specific cognitive skills that have been shown to be weak in reading-disabled children (Torgeson and Wagner, 1995; see also Box 6). Children who have the particular diagnostic pattern of cognitive weaknesses are selected for a matching kind of educational intervention that research has shown to be effective in increasing reading competence.

The assessment of cognitive correlates builds on a substantial body of research amassed over the last decade, especially on one kind of reading disability that appears to be a

BOX 6
Cognitive Assessment of Reading Disability: An Alternative Approach

Over the past 30 years, much of the research on reading disability has focused on trying to identify the "intrinsic cognitive limitations" that are the primary causes of this disability. In the absence of agreement about the nature of such cognitive processing, specific reading disability has been largely a category defined by exclusion. However, in recent years evidence has converged regarding one common subtype of reading disability, often called phonologically based reading disability; its underlying cognitive limitations are now well enough understood to allow the use of phonological assessment tasks as markers for this form of reading disability.

According to Torgeson, children who exhibit this form of reading disability have been shown to have weaknesses in at least three cognitive skills related to processing sounds in oral language: (1) *phonological awareness*: difficulty noticing, thinking about, or manipulating the individual sounds in words—assessed by tasks measuring sensitivity to individual phonemes within words and tasks measuring ability to blend separate phonemes together to form whole words (e.g., Bowey et al., 1992), (2) *phonological short-term memory:* difficulty remembering exact sequences of verbal information over brief periods of time—assessed by memory span tasks and tasks involving the repetition of complex nonwords or nonword phrases (e.g., Gathercole and Baddeley, 1993), and (3) *rapid automatic naming of verbal material:* difficulty accessing easily or rapidly phonological information that is stored in long-term memory—assessed by tasks that typically require naming, as rapidly as pos-

common kind—phonologically based reading disability (Rack et al., 1992; Share and Stanovich, 1995). Children with this reading disability have unique cognitive weaknesses of a specific nature only recently understood. Thus IQ tests, which measure a generic set of cognitive skills, were used to establish that reading failure was not the simple result of low intelligence. However, a presentation at the workshop suggests that the most common limitation in these children involves specific cognitive skills needed to process sounds in oral language (Liberman et al., 1989; Torgeson and Wagner, 1995), cognitive skills not specifically assessed by the more generalized IQ tests.

In recent years promising cognitive "markers" of disability have been identified and a battery of screening tests may have the potential to identify children who exhibit weak

sible, a series of 30 to 50 items printed on a page (e.g., Bowers and Swanson, 1991).

Evidence presented at the workshop suggests that these skills are highly stable individual difference variables that have concurrent and longitudinal relations with reading competence (see Share and Stanovich, 1995, for a recent review). Children with higher levels of phonological awareness, memory, and naming skills find it easier to learn to read in elementary school. Differences between disabled and average readers on these skills extend into adulthood. There is also evidence that the relation is not simply correlational but causal: individual differences in phonological awareness act as a primary cause of differences in the growth of word-reading skills (Hatcher et al., 1994; Wagner et al., 1995).

Studies discussed by Torgeson have shown that training in phonological awareness can have a positive effect on early reading skill (e.g., Ball and Blachman, 1991; Byrne and Fielding-Barnsley, 1993; Hatcher et al., 1994; Torgeson et al., 1992). These studies have indicated that combining training in phonological awareness with training in letter-sound correspondence may accelerate the rate of growth in reading skills. In addition, the studies suggest that children may require instruction designed specifically to help them generalize their new phonological awareness skills to the job of reading.

Based on a paper presented at the second workshop by Joseph Torgeson and Richard K. Wagner (1995).

phonological processing skills. This test battery, along with a reading achievement test, may allow early diagnosis of reading disability. Moreover, research on intervention methods suggests that some specific instructional strategies may work well with children so identified (e.g., Ball and Blachman, 1991; Byrne and Fielding-Barnsley, 1993; Hatcher et al., 1994; Torgeson et al., 1992).

Both of the lines of research described above have considerable appeal. Assessment is directly linked with instruction. Furthermore, children can be identified early, before school failure and low achievement become entrenched patterns. Finally, the assessments and instructional methods are grounded empirically. The next challenge lies in attempting to implement the findings of this and similarly promising research on a larger

scale. Such efforts will require educating teachers and administrators about alternative forms of assessment and instruction, and how they can influence instructional planning. This promise notwithstanding, 20 years of special education policies that encouraged the use of assessment for classification and eligibility decisions in a process quite separate from instructional planning will be hard to undo (see, for example, Klenk and Palincsar, 1994).

FUTURE RESEARCH

Throughout the workshop, participants identified the following promising areas for more and better research:

- *Effective treatment methods.* A better research base is needed regarding instructional methods that are effective with the lowest-achieving children. Good programming options are needed for children who are not succeeding in the general classroom. To the extent that we can bring more forms of effective educational methods into use, we can reduce the fear that the selection of a child for special education is tantamount to the assignment of that child to an inferior educational experience.

- *Assessment methods aligned with effective treatments.* Improved, alternative assessment methods are needed for making eligibility decisions and placing children in the most appropriate treatment for them. Assessments should provide information that can lead to adjustments in instruction and thereby improve student learning. Moreover, validity criteria for assessments need to include considerations of the consequences of their use in special education decision making—does the assessment information enable us to make better decisions about matching children to different educational treatments than would be made without such information? What are the social consequences of making inaccurate matches?

- *Analysis of factors in tests that contribute to differences in performance among racial, ethnic, gender, and disability groups.* Although a great deal of progress has been made in this area, more is needed. Among the factors that need to be better understood are motivation and the assumption of "maximum performance."

- *Research on the effects of various kinds of classroom environments on individual differences in learning.* In recent years research has deepened our understanding of the effects of situation and context on cognition and learning as well as documented the

effects of individual differences on performance in particular situations. "Ability and learning problems usually reside in the interface between person and situation—its mismatch—rather than in the person alone. This sort of deeper, richer theory of individual differences in cognitive learning, thinking, and reasoning in school situations is needed if more useful diagnostic assessments and improved instructional treatments are ever to be invented and developed" (Snow, 1995:35). In addition, research is needed into classroom practices that can help teachers set up learning environments that accommodate children with diverse learning needs.

- *Better understanding of what happens to children who encounter extreme academic difficulties in regular classrooms as they move through school systems, as well as the causes of the overrepresentation of minorities in this group.* Not much is known about the progress of low achievers over time. The long-term effects on children who are labeled as disabled or as having learning difficulties have long been of concern to educators and parents; more research is needed to understand the development of students with disabilities as they progress through school. In part the lack of research knowledge is because children labeled MR or SLD have often been held out of regular assessment programs and other large-scale educational research efforts, such as national longitudinal studies.

- *Deeper understanding of the malleability of intelligence and factors that contribute to enhancing a child's "ability to learn."* Although most researchers understand that "intelligence" is modifiable, many educators, policy makers, and the public still cling to the notion that human intelligence is largely innate and fixed. Research that explores the nature of environmental influences on cognitive development, as well as suggests instructional methods for teaching skills and strategies, can contribute to our understanding of malleability (see, e.g., Snow and Swanson, 1992).

Evidence and discussion at the two workshops sponsored by the Board on Testing and Assessment indicate that the usefulness of IQ tests in making special education decisions needs reevaluation. In the 20 years since IDEA was implemented, with a key role for testing and evaluation built into the eligibility process, research findings have grown—about tests, their utility in schools, and their role in instruction, as well as cognition, learning, and development—both normal and exceptional. Testing can play an important role in helping to match students with effective educational treatments. The time is ripe to focus on identifying such testing and instructional methods and moving them into educational practice.

REFERENCES

Ball, E.W., and B.A. Blachman
1991 Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly* 26:49-66.

Bowers, P.G., and L.B. Swanson
1991 Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology* 51:195-219.

Bowey, J.A., M.T. Cain, and S.M. Ryan
1992 A reading-level design study of phonological skills underlying fourth-grade children's word reading difficulties. *Child Development* 63:999-1011.

Brown, A.L., and J.C. Campione
1986 Psychological theory and the study of learning disabilities. *American Psychologist* 14(10):1059-1068.

Buss, W.G.
1995 Testing, Special Education, and Race Discrimination: Legal Framework and Policy Implications. Paper presented at a workshop of the Board on Testing and Assessment, La Jolla, CA, January 21, 1995. University of Iowa. (National Academy Press, 1996, forthcoming.)

Byrne, B., and R. Fielding-Barnsley
1993 Evaluation of a program to teach phonemic awareness to young children: a 1-year follow-up. *Journal of Educational Psychology* 85:104-111.

Ceci, S.
1990 *On Intelligence . . . More or Less*. Englewood Cliffs, N.J.: Prentice Hall.

Chapman, P.
1988 *Schools as Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890-1939*. New York: New York University Press.

Cole, N.S., and P.A. Moss
1989 Bias in test use. Pp. 201-219 in R.L. Linn, ed., *Educational Measurement*. New York: Macmillan.

Crawford et al. v. Honig
1988 C-89-0014 RFP United States District Court, Northern District of California.

Cronbach, L.
1975 Five decades of public controversy over mental testing. *American Psychologist* 30:1-13.
1990 *Essentials of Psychological Testing*. New York: Harper Collins.

Deno, S.L.
1985 Curriculum-based measurement: The emerging alternative. *Exceptional Children* 52:219-232.

Elliott, S.N.
1995 Performance Assessment of Students' Achievement: Research and Practice. Paper presented at a workshop of the Board on Testing and Assessment, Washington, DC, May 11, 1995. University of Wisconsin, Madison. (National Academy Press, 1996, forthcoming.)

Fancher, R.
1985 *The Intelligence Men: Makers of the IQ Controversy*. New York: Norton.

Feuer, M.
1994 Social policy, intelligence, and testing. Pp. 984-99 in R.J. Sternberg, ed., *Encyclopedia of Human Intelligence, Vol. 2*. New York: Macmillan.

Finn, J.D.
1982 Patterns in special education placement as revealed by OCR surveys. Pp. 322-381 in K.A. Heller, W.H. Holtzman, and S. Messick, eds., *Placing Children in Special Education: A Strategy for Equity*. Committee on Child Development Research and Public Policy, National Research Council. Washington, DC: National Academy Press.

Fletcher, J.M., D.J. Francis, S.E. Shaywitz, B.R. Foorman, K.K. Stuebing, and B.A. Shaywitz
1995 Diagnostic Utility of Intelligence Testing and the Discrepancy Model for Children with Learning Disabilities: Historical Perspectives and Current Research. Paper presented at a workshop of the Board on Testing and Assessment, Washington, DC, May 11, 1995. University of Texas Medical Center, Houston, Texas. (National Academy Press, 1996, forthcoming.)

Fletcher, J.M., S.E. Shaywitz, D.P. Shankweiler, L. Katz, I.Y. Liberman, A. Fowler, D.J. Francis, K.K. Stuebing, and B.A. Shaywitz
1994 Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology* 85:1-18.

Foorman, B.R., D.J. Francis, and J.M. Fletcher
1995 Growth of Phonological Processing Skills in Beginning Reading: The Lag Versus Deficit Model Revisited. Paper presented at the Society for Research on Child Development, Indianapolis, Indiana, March.

Francis, D.J., S.E. Shaywitz, K.K. Stuebing, B.A. Shaywitz, and J.M. Fletcher
1995 Developmental Lag Versus Deficit Models of Reading Disability: A Longitudinal Individual Growth Curves Analysis. Paper presented at the Society for Research on Child Development, Indianapolis, Indiana, March.

Fuchs, L.S.
1994 *Connecting Performance Assessment to Instruction*. Reston, VA: Council for Exceptional Children.
1995 Incorporating Curriculum-Based Measurement into the Eligibility Decision-Making Process: A Focus on Treatment Validity and Student Growth. Paper presented at a workshop of the Board on Testing and Assessment, Washington, DC, May 11, 1995. Peabody College of Vanderbilt University. (National Academy Press, 1996, forthcoming.)

Fuchs, D., and L.S. Fuchs
1995 What's "special" about special education? *Phi Delta Kappan* 76:522-530.

Gathercole, S.E., and A.D. Baddeley
1993 Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education* 8:259-272.

Gould, S.J.
1981 *The Mismeasure of Man*. New York: Norton.

Hatcher, P., C. Hulme, and A.W. Ellis
1994 Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development* 65:41-57.

Heller, K.A., W.H. Holtzman, and S. Messick, eds.
1982 *Placing Children in Special Education: A Strategy for Equity*. Committee on Child Development Research and Public Policy, National Research Council. Washington, DC: National Academy Press.

Herrnstein, R.
1971 IQ. *Atlantic Monthly* 228:43-64.

Holland, P.W., and H. Wainer, eds.
1993 *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Individuals with Disabilities Education Act
1991 20 U.S.C. Chapter 33.

Keogh, B.K., and D.L. MacMillan
1995 Exceptionality. In D. Berliner and R. Calfee, eds., *Handbook of Educational Psychology*. Washington, DC: American Psychological Association.

Klenk, L., and A.S. Palincsar
1994 Dynamic Classroom Observations: An Alternative Assessment in the Referral and Evaluation of Young Children Referred to Special Education. Paper presented at the National Reading Conference, San Diego, CA, November 1994, and distributed to participants at the workshop of the Board on Testing and Assessment, Washington, DC, May 11, 1995. University of Michigan.

Laosa, L.M.
1984 Social policies toward children of diverse ethnic, racial, and language groups in the United States. Pp. 1-109 in H.W. Stevenson and A.E. Siegel, eds., *Child Development Research and Social Policy, Vol. 1*. Chicago: University of Chicago Press.

Larry P. v. Riles
1972 *Larry P. v. Riles*. 343 F. Supp. 1306 (N.D. Cal.) (preliminary injunction). Affirmed 502.F 2d 963 (9th cir. 1974); 495 F. Supp. 926 (N.D. Cal. 1979) (decision on merits). Order modifying judgment, C-71-2270 RFP, Sept. 25, 1986.

Liberman, I.Y., D.P. Shankweiler, and A.M. Liberman
1989 The alphabetic principle and learning to read. In D.P. Shankweiler and I.Y. Liberman, eds., *Phonology and Reading Disability: Solving the Reading Puzzle* (IARLD Monograph Series). Ann Arbor: University of Michigan Press.

Lidz, C.S., ed.
1987 *Dynamic Assessment*. New York: Guilford Press.

Luckasson, R., ed.
1992 *Mental Retardation: Definition, Classification, and Systems of Support*. Washington, DC: American Association on Mental Retardation.

MacMillan, D.L.
1995 The Role of IQ in Eligibility and Placement Decisions for Children in Special Education. Paper presented at a workshop of the Board on Testing and Assessment, La Jolla, CA, January 21, 1995. University of California, Riverside. (National Academy Press, 1996, forthcoming.)

MacMillan, D., F.M. Gresham, G.N. Siperstein, and K.M. Bocian
1995 Definitions as Guides to Diagnostic Practices: An Examination of Public Schools' Eligibility and Placement Decisions in Cases Meeting the 1992 AAMR Definition of Mental Retardation. Unpublished paper. University of California, Riverside.

Marston, D., and D. Magnusson
1988 Curriculum-based assessment: District level implementation. Pp. 137-172 in J. Graden, J. Zins, and M. Curtis, eds., *Alternative Educational Delivery Systems: Enhancing Instructional Options For All Students*. Washington, DC: National Association of School Psychologists.

Marston, D., P. Mirkin, and S. Deno
1984 Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education* 18:109-117.

Mehan, H.
1995 The Institutional Uses of IQ Tests. Paper presented at a workshop of the Board on Testing and Assessment, La Jolla, CA, January 21, 1995. University of California, San Diego. (National Academy Press, 1996, forthcoming.)

BEST COPY AVAILABLE

Muehl, S., and E.R. Forell
1973 A follow-up study of disabled readers: Variables related to high school reading performance. *Reading Research Quarterly* 9:110-123.
PASE (Parents in Action on Special Education) v. Joseph P. Hannon
1980 *PASE (Parents in Action on Special Education) v. Joseph P. Hannon*. U.S. District Court, Northern District of Illinois, Eastern Division, No. 74 (3586).
Rack, J.P., M.J. Snowling, and R.K. Olson
1992 The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly* 27:29-53.
Reschly, D.J.
1995 IQ and Special Education: History, Current Status, and Alternatives. Paper presented at a workshop of the Board on Testing and Assessment, La Jolla, CA, January 21, 1995. Iowa State University. (National Academy Press, 1996, forthcoming.)
Reschly, D.J., and M.S. Wilson
1990 Cognitive processing vs. traditional intelligence: Diagnostic utility, intervention implications, and treatment validity. *School Psychology Review* 19:443-458.
Reschly, D.J., R.H. Kicklighter, and P. McKee
1988 Recent placement litigation. *School Psychology Review* 17:7-48.
Rutter M., and W. Yule
1975 The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry* 16:181-197.
Salvia, J., and J.E. Ysseldyke
1985 *Assessment in Special and Remedial Education*. Boston: Houghton Mifflin.
Sattler, J.M.
1982 *Assessment of Children's Intelligence and Special Abilities*. Boston: Allyn and Bacon.
Share, D.L., and K.E. Stanovich
1995 Cognitive processes in early reading development: A model of acquisition and individual differences. *Issues in Education: Contributions from Educational Psychology* 1:1-57.
Sherman, S.W., and N.M. Robinson, eds.
1982 *Ability Testing of Handicapped People: Dilemma for Government, Science, and the Public*. Committee on Ability Testing, National Research Council. Washington, DC: National Academy Press.
Shinn, M.R., ed.
1989 *Curriculum Based Measurement: Assessing Special Children*. New York: Guilford.
Shinn, M.R., G. Tindal, D. Spira, and D. Marston
1987 Practice of learning disabilities as social policy. *Learning Disability Quarterly* 10:17-28.
Smith, S.W.
1990 Individualized Education Programs (IEPs) in special education: From intent to acquiescence. *Exceptional Children* (September): 6-14.
Snow, R.E.
1995 Validity of IQ as a Measure of Cognitive Ability. Paper presented at a workshop of the Board on Testing and Assessment, La Jolla, CA, January 21, 1995. Stanford University. (National Academy Press, 1996, forthcoming.)
Snow, R.E., and I. Swanson
1992 Instructional pyschology: Aptitude, adaptation, and assessment. *Annual Review of Psychology* 43:583-626.
Snyderman, M., and S. Rothman
1988 *The IQ Controversy, the Media and Public Policy*. New Brunswick, NJ: Transaction Books.

Stanovich, K.E., and L.S. Siegel
1994 Phenotypic performance profiles of children with reading disabilities: A regression-based test of the phonological-core variable difference model. *Journal of Educational Psychology* 86:24-53.

Strag, G.A.
1972 Comparative behavioral ratings of parents with severe mentally retarded, special learning disability, and normal children. *Journal of Learning Disabilities* 5:52-56.

Torgeson, J.K., and R.K. Wagner
1995 Alternative Diagnostic Approaches for Specific Developmental Reading Disabilities. Paper presented at a workshop of the Board on Testing and Assessment, Washington, DC, May 11, 1995. Florida State University. (National Academy Press, 1996, forthcoming.)

Torgeson, J.K., S. Morgan, and C. Davis
1992 The effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology* 84:364-370.

U.S. Department of Education
1994 *Sixteenth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: Office of Special Education Programs.

Wagner, R.K., J.K. Torgeson, C.A. Rashotte
1995 The Development of Reading-related Phonological Processes: A Five Year Longitudinal Study. Paper presented at the Society of Research in Child Development, Indianapolis, Indiana, April.

Wigdor, A., and W. Garner, eds.
1982 *Ability Testing: Uses, Consequences, and Controversies*. Committee on Ability Testing, National Research Council. Washington, DC: National Academy Press.

Willingham, W.W., M. Ragosta, R.E. Bennett, H. Braun, D.A. Rock, and D.E. Powers
1988 *Testing Handicapped People*. Needham Heights, MA: Allyn and Bacon.

Witt, J.C., and F.M. Gresham
1985 Review of the Wechsler Intelligence Scale for Children Revised. Pp. 1716-1719 in J. Mitchell, ed., *Ninth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute.

APPENDIX

WORKSHOP AGENDAS AND PARTICIPANTS

IQ TESTING AND EDUCATIONAL DECISION MAKING

WORKSHOP #1

Board on Testing and Assessment

January 21, 1995

Sheldon H. White, Chair

Harvard University

AGENDA

10:00 am Welcome and Introductions
Richard Atkinson, Chair, BOTA
Sheldon White

10:30 IQ Testing and Special Education: The Legal and Policy Context
William Buss, University of Iowa
Discussant: William Taylor (BOTA)

12:00 Lunch

12:30 pm Validity of IQ as a Measure of Cognitive Ability
Richard Snow, Stanford University
Discussant: James Greeno, Stanford University

2:00 IQ Testing and Special Education: Diagnosis, Classification, and Placement
Daniel Reschly, Iowa State University
Donald MacMillan, University of California, Riverside
Discussants: Cynthia Hudley-Paul,
University of California, Santa Barbara
Milton Budoff,
Research Institute for Educational Problems

4:00 Coffee Break

4:15 IQ and School Decision Making: Evidence from the Field
Hugh Mehan, University of California, San Diego
Discussants: Richard Durán,
University of California, Santa Barbara
Stephen Elliott,
University of Wisconsin, Madison

5:45 Break

6:00 Dinner

7:15 Summary: David Berliner (BOTA)

7:45 Conclusions and Next Steps: Sheldon White

PARTICIPANTS

RICHARD ATKINSON, University of California, San Diego

LAURIE BASSI, Graduate Public Policy Program, Georgetown University

DAVID C. BERLINER, College of Education, Arizona State University

SUSAN BOWERS, Office for Civil Rights, U.S. Department of Education

JEFFERY BRADEN, Department of Educational Psychology, University of Wisconsin, Madison

MILTON BUDOFF, Research Institute for Educational Problems, Cambridge, MA

WILLIAM BUSS, College of Law, University of Iowa

GARY CARTER, Office of Policy and Research, U.S. Department of Labor

ARTHUR COLEMAN, Office for Civil Rights, U.S. Department of Education

LOUIS DANIELSON, Special Education, U.S. Department of Education

RICHARD DURÁN, Graduate School of Education, University of California, Santa Barbara

STEPHEN ELLIOTT, Department of Educational Psychology, University of Wisconsin, Madison

MICHAEL FEUER, National Research Council, Washington, DC

JAMES GREENO, Department of Education, Stanford University

JANET HANSEN, National Research Council, Washington, DC

CYNTHIA HUDLEY-PAUL, Graduate School of Education, University of California, Santa Barbara

SYLVIA JOHNSON, School of Education, Howard University

CARL KAESTLE, School of Education, University of Chicago

NADINE LAMBERT, School of Education, University of California, Berkeley

LUIS LAOSA, Educational Testing Service, Princeton, NJ

RENÉE LERCHE, Ford Motor Company, Dearborn, MI

JESSICA LEVIN, Office of the Under Secretary, U.S. Department of Education

NORMAN LEWIS, Ford Motor Company, Dearborn, MI

ROBERT LINN, School of Education, University of Colorado, Boulder

DONALD MACMILLAN, School of Education, University of California, Riverside

SHIRLEY MCGUIRE, Department of Psychology, University of California, San Diego

HUGH MEHAN, Teachers Education Program, University of California, San Diego

PATRICIA MORISON, National Research Council, Washington, DC

MARY QUENETTE, Defense Manpower Data Center, U.S. Department of Defense

DANIEL RESCHLY, Department of Psychology, Iowa State University
JOLETA REYNOLDS, Special Education Programs, U.S. Department of Education
NEAL SCHMITT, Department of Psychology, Michigan State University
RICHARD SHAVELSON, School of Education, Stanford University
SUZANNE SHERIDAN, Office of the General Counsel, U.S. Department of Education
RICHARD SNOW, School of Education, Stanford University
WILLIAM TAYLOR, Attorney at Law, Washington, DC
EWART THOMAS, Department of Psychology, Stanford University
SHELDON WHITE, Department of Psychology, Harvard University

IQ TESTING AND EDUCATIONAL DECISION MAKING

WORKSHOP #2

Board on Testing and Assessment

May 11, 1995

Sheldon H. White, Chair

Harvard University

AGENDA

9:00 am	Welcome and Introduction Sheldon White
9:30	Diagnostic Use of IQ Tests: Historical Perspectives and Current Research Jack Fletcher, University of Texas Medical Center - Houston
	Alternative Approaches to Assessment:
10:15	Diagnosis and Intervention with Reading Disabilities Joseph K. Torgeson, Florida State University
11:00	Break
11:15	Curriculum Based Measurement Lynn Fuchs, Peabody College/Vanderbilt University
12:00	Lunch

1:00 pm Performance Assessment
Stephen Elliott, University of Wisconsin, Madison

1:45 Dynamic Classroom Observations
Annmarie Palincsar, University of Michigan

2:30 Break

Synthesis and Reaction to Alternative Approaches:

2:45 Educational Perspectives
Discussants: Donald MacMillan,
University of California, Riverside

Deborah Speece,
University of Maryland

3:45 Legal Perspectives
Discussants: William Buss,
University of Iowa

William Taylor,
BOTA

4:30 Wrap-up and Closing Remarks: Sheldon White

PARTICIPANTS

JOHN ALGER, Office for Civil Rights, U.S. Department of Education

LISA BATTALIA, Special Education Programs, U.S. Department of Education

DAVID BERKOWITZ, Office for Civil Rights, U.S. Department of Education

CANDACE BOS, Department of Special Education and Rehabilitation, University of Arizona

SUSAN BOWERS, Office for Civil Rights, U.S. Department of Education

WILLIAM BUSS, College of Law, University of Iowa

ARTHUR COLEMAN, Office for Civil Rights, U.S. Department of Education

LOUIS DANIELSON, Special Education Programs, U.S. Department of Education

STEPHEN ELLIOTT, Department of Educational Psychology, University of Wisconsin, Madison

MICHAEL FEUER, National Research Council, Washington, DC

JACK FLETCHER, University of Texas Medical Center, Houston, TX

LYNN FUCHS, Department of Special Education, Peabody College, Vanderbilt University

ANNE HOOGSTRATEN, Office for Civil Rights, U.S. Department of Education

LUIS LAOSA, Educational Testing Service, Princeton, NJ

ALAN LESGOLD, Learning, Research, and Development Center, University of Pittsburgh

SYLVIA JOHNSON, School of Education, Howard University

JESSICA LEVIN, Office of the Under Secretary, U.S. Department of Education

ROBERT LINN, School of Education, University of Colorado, Boulder

REID LYON, National Institutes of Health, Bethesda, MD

DONALD MACMILLAN, School of Education, University of California, Riverside

DAVID MALOUF, Special Education Programs, U.S. Department of Education

PATRICIA MORISON, National Research Council, Washington, DC

JACQUES NORMAND, National Research Council, Washington, DC

ANNMARIE PALINCSAR, School of Education, University of Michigan, Ann Arbor

JEAN PEELEN, Office for Civil Rights, U.S. Department of Education

DANIEL RESCHLY, Department of Psychology, Iowa State University

JOLETA REYNOLDS, Special Education Programs, U.S. Department of Education

HENRY RIECKEN, Washington, DC

STACEY ROSEBERRY, Office for Civil Rights, U.S. Department of Education

DEBORAH SPEECE, Department of Special Education, University of Maryland, College Park

BENNETT SHAYWITZ, Department of Pediatrics, Yale University School of Medicine
SALLY SHAYWITZ, Department of Pediatrics, Yale University School of Medicine
SUZANNE SHERIDAN, Office of the General Counsel, U.S. Department of Education
WILLIAM TAYLOR, Attorney at Law, Washington, DC
JOSEPH K. TORGESON, Department of Psychology, Florida State University
SHELDON WHITE, Department of Psychology, Harvard University
ALEXANDRA WIGDOR, National Research Council, Washington, DC

About the Board on Testing and Assessment

The Board on Testing and Assessment was established in 1994 with support from the United States Departments of Defense, Education, and Labor. Its principal objectives are to aid policy makers in the clarification of the purposes of testing and assessment and to help them evaluate the uses of tests, alternative assessments, and other indicators commonly used as tools of public policy. The board brings to bear the knowledge and tools of the social and behavioral sciences and provides an analytical base for the examination of difficult issues in the measurement and evaluation of achievement in education, the workplace, and other settings. The board is a long-term activity of the National Research Council designed to be responsive to evolving challenges that face schooling, work, and the recruitment of human competencies. Some specific functions of the board include analyzing and advising on the science of testing and assessment, providing a forum for discussion of issues about which to discuss the effects of planned testing and assessment programs, helping government agencies coordinate their policies and practices, and providing technical and policy problems in testing and assessment.

Members 1995-1996

Richard J. Shavelson, Chair, Stanford University
Laura J. Bassi, Vice Chair, American Society for Training and Development
Robert L. Linn, Vice Chair, University of Colorado, Boulder
Richard C. Atkinson, University of California, Oakland
Isabel G. Barnes, Potomac Electric Power Company
David C. Berliner, Arizona State University, Tempe
Paul J. Black, King's College London
Richard E. Elmore, Harvard University
Arthur S. Goldberger, University of Wisconsin, Madison
Edmund W. Gordon, City University of New York
Paul W. Holland, University of California, Berkeley